. 23**MAT112**

Mathematics for Intelligent Systems – 2

Practice Sheet – 3

Projection, Least Squares, Regression lines

1. **Pseudoinverse of an m×n matrix, A**

Every matrix $A_{m\times n}$ has a pseudoinverse $\mathbf{A^+}$ or **pinv(A)**

- If m = n = rank of A, then A is invertible and $\mathbf{A^+ = A^{-1}}$
- If rank of A = n, (all columns of A are independent), then $\mathbf{A^+ = (A^TA)^{-1}A^T}$ **= Left Inverse of A,** and $\mathbf{A^+A = I_n}$.
- If rank of A = m, (all rows of A are independent), then $\mathbf{A^+ = A^T (AA^T)^{-1}}$ **= Right Inverse of A,** and $\mathbf{AA^+ = I_m}$.

- Pseudoinverse $\mathbf{A^+}$ of any matrix A can be obtained in MATLAB using the command 'pinv(A)'.

a. Generate a 9×9 random integer matrix and verify that inv(A) = pinv(A).

b. Generate a 5×4 random integer matrix and verify that pinv(A) is same as the left inverse of A.

c. Generate a 3×7 random integer matrix and verify that pinv(A) is same as the right inverse of A.

2. Projection of a vector $\boldsymbol{v}$ on a subspace S, which is a column space of matrix A(with all independent columns) is $\mathbf{p} = A(A^TA)^{-1}A^T\mathbf{v}$.

If some columns in A are dependent (i.e., if $\text{Rank}(A_{m\times n}) < n$), then the projection of a vector $\boldsymbol{v}$ on a subspace S, which is a column space of matrix A is **p** = A*pinv(A)***v**, where pinv(A) is the pseudo inverse of A.

a. Find the projection of $v = \begin{pmatrix} -3 \\ -3 \\ 8 \\ 9 \end{pmatrix}$ on the space spanned by $\begin{pmatrix} 3 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 2 \\ -1 \end{pmatrix}$.

b. Find the projection of $u = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$ on the space spanned by $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$.

c. Find the projection of $u = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$ on the column space of $= \begin{pmatrix} 1 & 1 & 2 & 4 \\ 1 & 2 & 3 & 6 \\ 0 & 1 & 1 & 2 \end{pmatrix}$.

Consider the matrices given below: A, B, C, D and E.

(a) Which of the matrices have inverse? Find if it exists.

(b) Which of the matrices will have a left inverse? Find if it exists.

(c) Which of the matrices will have a right inverse? Find if it exists.

(d) Find pseudo inverse of all these matrices using the command 'pinv'. Compare the answers with the answers obtained in (a), (b), (c) and (d).

$$A=\begin{bmatrix}1 & 4 & 5 & 6 & 9\\ 3 & -2 & 1 & 4 & -1\\ -1 & 0 & -1 & -2 & -1\\ 2 & 3 & 5 & 7 & 8\end{bmatrix},\quad B=\begin{bmatrix}3 & 4\\ 1 & 8\end{bmatrix},\quad C=\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 0 & 4 & 2\end{bmatrix},$$

$$D=\begin{bmatrix}1 & 2 & 0 & 1\\ 0 & 1 & 1 & 0\\ 2 & 4 & 0 & 2\end{bmatrix},\qquad E=\begin{bmatrix}0 & 1 & 4 & 1\\ 0 & 0 & 8 & 1\\ 0 & 1 & 4 & 0\\ 1 & 0 & 8 & 0\\ 0 & 1 & 1 & 0\\ 1 & 0 & 1 & 0\end{bmatrix}$$

.

3. Least square solution of a linear system, $Ax = b$ is:

$\hat{x} = (A^TA)^{-1}A^Tb$, if all the column vectors in A are independent (i.e. if A^TA is invertible).

$\hat{x} = \text{pinv}(A)*b$, if the column vectors in A are not all independent.

- Find the least square solution for the following systems:
  a. x + y = 10; 2x + y = 10; 3x + y = 16; 4x + y = 27; 5x + y = 23.
  b. 2x – 3y + z = 1; x + 3y – 4z = -4; 3x – 2y – z = -1; 5x – y – 4z = -4.
  c. 2x + y = 20; 2x - 3y = 4; 3x + y = 28; 4x + y = 36;

3. Fit a parabola (quadratic function) with (-2, 0), (-1,0), (0,1), (1, 0), and (2, 0).

4. A middle-aged man was stretched on a rack to lengths L = 5, 6, and 7 feet under applied forces of F = 1, 2, and 4 tons. Assuming Hookes law L = a + bF , find his normal length a by leastsquares.

5. Find the best straight-line t (least squares) to the measurements: b = 4 at t = -2, b = 3 at t = -1, b = 1 at t = 0 and b = 0 at t = 2. Then find the projection of b = (4; 3; 1; 0) onto the column space of A = $\begin{bmatrix}1 & -2\\ 1 & -1\\ 1 & 0\\ 1 & 2\end{bmatrix}$.

6. Fit a best straight line passing through (-3, 4),(-2,2),(-1,0),(0,0),(1,-2),(2,-2) and (3,-3).

7. Consider the following data:

xi	yi
0	-2.87
1	1.93
2	6.33
3	12.48
4	15.98
5	21.77
6	25.89
7	32.01
8	36.21
9	43.65
10	47.09
11	50.97
12	56.57
13	62.25
14	65.77
15	71.91
16	77.85
17	80.66
18	89.34
19	93.62
20	99.01
21	101.29
22	108.11
23	112.21
24	116.77
25	123.01

a. Find the equation of a regression line for the following data using excel.

b. Enter the data in an excel sheet, and then find the regression line using MATLAB. Also plot the regression line and the data points in the same figure.

- Refer to the class notes for this.

Fitting of a regression line

	A	B	C	D	E
1		x	y	xiyi	xi squared
2		-10	-130.771	1307.711	100
3		-9	-124.911	1124.2	81
4		-8	-188.656	1509.25	64
5		-7	-87.4707	612.2951	49
6		-6	3.996329	-23.978	36
7		-5	-133.754	668.7695	25
8		-4	-69.7174	278.8696	16
9		-3	21.11265	-63.338	9
10		-2	5.479887	-10.9598	4
11		-1	0.237035	-0.23704	1
12		0	41.39672	0	0
13		1	42.67804	42.67804	1
14		2	155.2094	310.4189	4
15		3	109.763	329.289	9
16		4	172.5026	690.0104	16
17		5	155.0395	775.1975	25
18		6	203.447	1220.682	36
19		7	229.9432	1609.603	49
20		8	238.0019	1904.016	64
21		9	260.1743	2341.569	81
22		10	245.2808	2452.808	100
23	Sums	0	1148.982	17078.85	770

	H	I	J	K
5	n=21			
7	B Matrix			b vector
9	770	0		17078.85
10	0	21		1148.982

Since $\sum_i x_i = 0$ in our Example

$$m = \frac{\sum_i x_i y_i}{\sum_i x_i^2},$$

$$c = \frac{\sum_i y_i}{n}$$

m=	22.18033
c=	54.71343

	A	B	C
1	x	y	
2	-10	-130.771	
3	-9	-124.911	
4	-8	-188.656	
5	-7	-87.4707	
6	-6	3.996329	
7	-5	-133.754	
8	-4	-69.7174	
9	-3	21.11265	
10	-2	5.479887	
11	-1	0.237035	
12	0	41.39672	
13	1	42.67804	
14	2	155.2094	
15	3	109.763	
16	4	172.5026	
17	5	155.0395	
18	6	203.447	
19	7	229.9432	
20	8	238.0019	
21	9	260.1743	
22	10	245.2808	

Write a MATLAB code to compute m, c using the matrix equation
Data should be read from an excel file

Steps

1. Create an excel file with your data as in the figure. Name it as xlformatlab.xlsx
2. Write the matlab code as follows

```
[B,text]=xlsread('xlformatlab.xlsx');
x=B(:,1);
y=B(:,2);
n=length(x);
ATA=[sum(x.^2),sum(x);sum(x),n]
ATb=[sum(x.*y);sum(y)]
m_and_c= inv(ATA)*ATb
```

B will contain your numeric data in a matrix format. 'text' you can discard

	A	B	C
1	x	y	
2	-10	-130.771	
3	-9	-124.911	
4	-8	-188.656	
5	-7	-87.4707	
6	-6	3.996329	
7	-5	-133.754	
8	-4	-69.7174	
9	-3	21.11265	
10	-2	5.479887	
11	-1	0.237035	
12	0	41.39672	
13	1	42.67804	
14	2	155.2094	
15	3	109.763	
16	4	172.5026	
17	5	155.0395	
18	6	203.447	
19	7	229.9432	
20	8	238.0019	
21	9	260.1743	
22	10	245.2808	

Write a MATLAB code to find the regression line and also plot it along with the given data points

- Create an excel file with your data as in the figure. Name it as xlformatlabReg.xls and use the code below:

```
[B,text]=xlsread('xlformatlabReg.xlsx');
x=B(:,1);
y=B(:,2);
n=length(x);
A=[x,ones(n,1)];
ATA=[sum(x.^2),sum(x);sum(x),n]
ATb=[sum(x.*y);sum(y)]
m_and_c= inv(ATA)*ATb
y_projected=A*inv(ATA)*ATb;
plot(x,y,'*')
hold on
plot(x,y_projected)
```

8. a. Generate a data for regression with a fixed m=9, c=5 for x = -25,-24, -23,…,24,25 and introducing a noise with standard deviation 3.

b. Using the data generated above, re-estimate the values of m and c. Also find the regression line and plot it in MATLAB.

- Refer to the class notes for this.

Linear Regression : Generation of data and fitting

```
m=7;
c=3;
x=-5:0.1:5; % a row vector x
n=length(x);
y=m*x+c;
noise_stdev=6;
       % noise standard deviation
noise=noise_stdev*randn(1,n);
yn=y+noise;
plot(x,yn,'*');
       % plot of generated points
hold on
  % fitting a line or re-estimating m and c
  % create matrix A with x and ones
A=[x' ones(n,1)];
yn=yn'; % make yn a column vector
m_and_c=pinv(A)*yn; % estimate m and c
y_projected=A*pinv(A)*yn; % y estimated
plot(x,y_projected);
m_and_c
```